

明 細 書

会話支援装置及び会話支援方法

技術分野

- [0001] 本発明は、例えば障害者との会話を支援する会話支援装置及び会話支援方法に関する。

背景技術

- [0002] 従来から、障害者との会話、あるいは外国人との会話を支援する技術が提案されている。例えば、ビデオカメラで撮影した話者の唇の動き等の画像を画像解析し、これに対応した文字情報や音声情報に変換して表示部やスピーカ等に出力するという技術がある(例えば、日本国特許庁が発行する公開特許公報―特開2004-15250号公報(段落[0017]、図1)参照。))。
- [0003] また、話者の話す言語を所定の言語に翻訳して、その翻訳された言語をモニター等に文字列で呈示するという技術もある(例えば、同じく公開特許公報―特開2003-345379号公報(段落[0013]、図1)参照。))。

発明の開示

発明が解決しようとする課題

- [0004] 実際に、話者と聴者とが向き合って話す場合に、例えば上記特許文献1の装置を用いた場合に、さらに効率よく画像処理や音声処理等がなされることが望まれる。特に、特許文献1には、具体的な読唇処理の手法は記載されていない。したがって、このままでは、この特許文献1に係る装置は実現不可能である。
- [0005] また、例えば、特許文献1の装置では、画像解析の制御キーが話者によって押されることにより、カメラが活性状態となり、活性状態となってから画像解析処理が開始されるようになっている。したがって、話者は話すごとに当該制御キーを押さなければならず不便である。
- [0006] 以上のような事情に鑑み、本発明の目的は、効率よくスムーズに会話することができる会話支援装置及び会話支援方法を提供することにある。

課題を解決するための手段

- [0007] 上記目的を達成するため、本発明に係る会話支援装置は、第1の利用者及び第2の利用者を撮像する手段と、前記撮像された前記第1の利用者の第1の画像情報を前記第2の利用者が理解可能な第1の情報に変換する第1の変換手段と、前記撮像された前記第2の利用者の第2の画像情報を前記第1の利用者が理解可能な第2の情報に変換する第2の変換手段と、第1の表示面を有し、前記変換された前記第1の情報を前記第1の表示面に表示する第1の表示手段と、前記第1の表示面とは異なる角度に配置可能な第2の表示面を有し、前記変換された前記第2の情報を前記第2の表示面に表示する第2の表示手段とを具備する。
- [0008] 本発明では、第1の表示面と第2の表示面とが異なる角度に配置可能であるので、第1の利用者が第2の表示面を見ながら、また、第2の利用者が第1の表示面を見ながら効率よくスムーズに会話することができる。また、第1の表示面と第2の表示面とが備えられることにより、例えば第2及び第1の利用者が対面して会話することができる。
- [0009] 本発明において、第1の利用者は、健常者または聴覚障害者である。第1の利用者が健常者である場合、第2の利用者は聴覚障害者である。第1の利用者が聴覚障害者である場合、第2の利用者は健常者または聴覚障害者である。
- [0010] 一方、本発明に係る会話支援装置は、外国人との会話を支援することも可能である。この場合、例えば、第1の利用者が第1の言語を話す人である場合、第2の利用者が第1の言語とは異なる第2の言語を話す人である。
- [0011] 本発明の一の形態によれば、前記第1の利用者の声音を集音する手段と、前記集音された声音を前記第1の情報に変換する第3の変換手段とをさらに具備する。これにより、例えば言葉を話すことができる話者であれば、撮像した画像情報を用いなくとも、当該話者の声音の情報のみを用いて第1の情報に変換処理することができる。あるいは、画像解析処理と声音解析処理と両方行われるようにしてもよい。これにより、第1の利用者が第2の利用者に伝えようとしている内容の認識精度を向上させることができる。本発明では、同様に第2の利用者の声音を集音する手段も設けるようにしてもよい。
- [0012] 本発明においては、例えば、集音された第1の利用者の声音の音声パターンと、第

1の情報のパターンとを対応付けて記憶する手段を用意しておくようにすればよい。これにより、第3の変換手段が、当該音声情報を解析して上記音声パターンを決定し、その音声パターンに対応する第1の情報のパターンを抽出することによって変換処理が可能となる。第2の利用者の声音も同様に処理することができる。

[0013] 本発明の一の形態によれば、前記第1の変換手段は、前記第1の画像情報を、前記第1の情報として文字情報及び手話情報のうちいずれか一方に変換する手段を有する。これにより、例えば第2の利用者が聴覚障害者である場合に、当該第2の利用者は文字情報や手話情報を第1の表示面で見ることによって会話が可能となる。第2の変換手段も同様に、第2の利用者の画像情報を第2の情報として文字情報及び手話情報のうちいずれか一方に変換する手段を有していてもよい。これにより、第1の利用者が聴覚障害者であってもよく、両者が聴覚障害者である場合であっても効率よく会話することができる。

[0014] 本発明の一の形態によれば、前記第1の変換手段は、前記第1の画像情報を、前記第1の情報として音声情報に変換する手段を有し、当該会話支援装置は、前記変換された音声情報を出力する手段をさらに具備する。これにより、例えば第1の利用者が健常者である場合はもちろんのこと、第1の利用者が聴覚障害者である場合に、当該聴覚障害者と、視覚障害者(第2の利用者)との会話を成立させることができる。また、第2の変換手段も同様に、第2の利用者の画像情報を、第2の情報として音声情報に変換する手段を有し、当該会話支援装置は、変換された音声情報を出力する手段をさらに具備してもよい。

[0015] 本発明の一の形態によれば、前記第1の表示手段は、前記第1の表示面に前記第2の情報を表示する手段を有する。これにより、第2の利用者は、第1の表示面を見て第2の情報を自ら確認することができる。例えば第2の変換手段によって変換ミスがないかどうかを確認することができる。また、第2の表示手段も、前記第2の表示面に前記第1の情報を表示する手段を有していてもよい。特に、本発明の一の形態によれば、前記第1の表示手段は、前記第2の表示手段が前記第2の情報を表示するタイミングとほぼ同期して該第2の情報を表示する手段を有していれば、会話がよりスムーズに進行する。

- [0016] 本発明の一の形態によれば、前記第1の変換手段が、前記第1の画像情報を文字情報及び手話情報に変換する場合、前記第1の表示手段は、前記第1の変換手段で変換された前記手話情報を前記第1の表示面に表示し、前記第2の表示手段は、前記第1の表示手段が前記手話情報を表示するタイミングとほぼ同期して、前記第1の変換手段で変換された前記文字情報を表示する手段を有する。本発明では、第1の利用者が第2の利用者に伝えようとする内容を、第2の利用者が見る第1の表示面に手話情報として表示され、第1の利用者が見る第2の表示面には文字情報で表示される。これにより、第1の利用者は、自分が伝えようとしている内容を画面を見ながら確認できる。特に、変換ミス等がないかを確認することができる。
- [0017] 本発明の一の形態によれば、前記第1の利用者の声音の音圧レベルまたは騒音レベルを計測する手段と、前記音圧レベルまたは騒音レベルの閾値を設定する手段と、前記測定された音圧レベルまたは騒音レベルが前記閾値以上である場合に、前記第3の変換手段による変換処理を実行する手段とをさらに具備する。これにより、例えば第1の利用者が閾値以上の音圧レベルまたは騒音レベルで話を開始すれば、自動的に音声認識を開始することができる。したがって、従来のように、話し始めるたびに画像解析や音声認識の制御キーのようなものを押す必要がなくなり、会話をスムーズにすることができる。また、本発明では、上記第3の変換手段に代えて、または、第3の変換手段とともに、第1の変換手段による変換処理を実行するようにしてもよい。
- [0018] 本発明の一の形態によれば、人が発声したときの該人の口唇の複数の画像パターンと、該人が発声したときの音に対応する前記文字情報とを対応付けて記憶する手段をさらに具備し、前記撮像手段は、前記第1の画像情報として第1の利用者の口唇の画像を撮像し、前記第1の変換手段は、前記撮像された口唇画像情報と、前記各画像パターンとを比較する手段と、比較の結果、該各画像パターンのうち前記撮像された口唇画像情報に対応する一の画像パターンを選択する手段と、選択された前記一の画像パターンに対応する前記文字情報を抽出する手段とを有する。人とは、第1の利用者自身であってもよいし、第1の利用者とは別人であってもよい。また、第2の利用者とは別人であってもよい。以下、同様である。本発明においては、第1の変換手段は、いわゆるパターンマッチングにより撮像された口唇画像情報に対応する画

像パターンを選択する。もちろん第2の変換手段が、このような第1の変換手段と同様の手段を有するようにしてもよい。

[0019] 本発明の一の形態によれば、前記記憶手段は、複数の人ごとに、前記画像パターンと前記文字情報とを対応付けて記憶する手段と、前記各人の口唇の形状パターンを記憶する手段とを有し、当該会話支援装置は、前記撮像された口唇画像情報と、前記各形状パターンを比較する手段と、比較の結果、前記各形状パターンのうち、前記撮像された口唇画像情報に最も類似する一の形状パターンを選択する手段とをさらに具備する。本発明では、最も類似する一の形状パターンを選択したら、第1及び第2の利用者の会話を開始し、選択された一の形状パターンを基に上記パターンマッチングを行うことができる。これにより、より高精度に利用者の画像認識を行うことができ、また、より高精度に変化処理を行うことができる。したがって変換処理の変換ミス等の発生率を低減することができる。

[0020] 本発明の一の形態によれば、前記人が発声したときの該人の口唇の複数の画像パターンと、該人が発声したときの音に対応する前記手話情報とを対応付けて記憶する手段をさらに具備し、前記撮像手段は、前記第1の画像情報として第1の利用者の口唇の画像を撮像し、前記第1の変換手段は、前記撮像された口唇画像情報と、前記各画像パターンとを比較する手段と、比較の結果、前記撮像された口唇画像情報に対応する該各画像パターンのうちの画像パターンを選択する手段と、選択された前記一の画像パターンに対応する前記手話情報を抽出する手段とを有する。もちろん第2の変換手段が、このような第1の変換手段と同様の手段を有するようにしてもよい。さらに、本発明では、前記記憶手段は、複数の人ごとに、前記画像パターンと前記手話情報とを対応付けて記憶する手段と、前記各人の口唇の形状パターンを記憶する手段とを有し、当該会話支援装置は、前記撮像された口唇画像情報と、前記各形状パターンを比較する手段と、比較の結果、前記各形状パターンのうち、前記撮像された口唇画像情報に最も類似する一の形状パターンを選択する手段とをさらに具備するようにしてもよい。

[0021] 本発明の一の形態によれば、前記人が発声したときの該人の口唇の複数の画像パターンと、該人が発声したときの音に対応する前記音声情報とを対応付けて記憶す

る手段をさらに具備し、前記撮像手段は、前記第1の画像情報として第1の利用者の口唇の画像を撮像し、前記第1の変換手段は、前記撮像された口唇画像情報と、前記各画像パターンとを比較する手段と、比較の結果、前記撮像された口唇画像情報に対応する該各画像パターンのうちの画像パターンを選択する手段と、選択された前記一の画像パターンに対応する前記音声情報を抽出する手段とを有する。もちろん第2の変換手段が、このような第1の変換手段と同様の手段を有するようにしてもよい。さらに、本発明では、前記記憶手段は、複数の人ごとに、前記画像パターンと前記手話情報とを対応付けて記憶する手段と、前記各人の口唇の形状パターンを記憶する手段とを有し、当該会話支援装置は、前記撮像された口唇画像情報と、前記各形状パターンを比較する手段と、比較の結果、前記各形状パターンのうち、前記撮像された口唇画像情報に最も類似する一の形状パターンを選択する手段とをさらに具備するようにしてもよい。

[0022] 本発明の一の形態によれば、前記第1の表示手段は、第1の端部を有し前記第1の表示面が配置される第1の部分とを有し、前記第2の表示手段は、第2の端部を有し前記第2の表示面が配置され、前記第1の表示面に対し前記第2の表示面をねじることができるように、かつ、前記第1の部分に対し折りたたみできるように、前記第1の端部と前記第2の端部とが接続された第2の部分とを有する。このような構成により、第1の利用者と第2の利用者との配置が問われない。

[0023] 本発明の一の形態によれば、前記第1の情報を記憶する手段と、前記第1の情報を記憶する手段と、前記記憶された第1の情報を前記第1の表示面に再生する手段と、前記記憶された第1の情報をトリックプレイ操作する手段とをさらに具備する。トリックプレイとは、早送り、巻戻し、スロー再生、2倍速再生等のことであり、通常の画像の再生とは異なる表示手法をいう。これにより、例えば第2の利用者が見逃した情報を巻戻しする等して再生することができる。また、会話支援装置は、第2の情報を記憶する手段と、第2の表示手段により前記第1の情報をトリックプレイで表示させるために、記憶された第2の情報にトリックプレイ操作を行う手段とをさらに具備するようにしてもよい。

[0024] 本発明に係る会話支援方法は、第1の利用者及び第2の利用者を撮像するステッ

プと、前記撮像された前記第1の利用者の画像情報を前記第2の利用者が理解可能な第1の情報に変換するステップと、前記撮像された前記第2の利用者の画像情報を前記第1の利用者が理解可能な第2の情報に変換するステップと、前記変換された前記第1の情報を前記第1の表示面に表示するステップと、前記変換された前記第2の情報を、前記第1の表示面とは異なる角度に配置可能な第2の表示面を表示するステップとを具備する。

- [0025] 本発明では、第1の表示面と第2の表示面とが異なる角度に配置可能であるので、第1の利用者が第2の表示面を見ながら、また、第2の利用者が第1の表示面を見ながら効率よくスムーズに会話することができる。

発明の効果

- [0026] 以上のように、本発明によれば、効率よくスムーズに会話することができる。

発明を実施するための最良の形態

- [0027] 以下、本発明の実施の形態を図面に基づき説明する。

- [0028] 図1は、本発明の一実施の形態に係る会話支援装置を示す斜視図である。図2は、この会話支援装置1の構成を示すブロック図である。

- [0029] 会話支援装置1は、第1の本体20と、第2の本体30とで構成されている。会話支援装置1は、例えばPDA(Personal Digital Assistance)程度の大きさを有する。第1の本体20には、第1の表示部22が設けられ、第2の本体30にも同様に第2の表示部32が設けられている。第1の表示部22は、例えば液晶や、有機EL(Electro-Luminescence)等が用いられる。第2の表示部32も第1の表示部22の構成と同様である。第1の本体20の一端部2b及び第2の本体30の一端部3bにはカメラ21及び31それぞれが設けられている。カメラ21及び31はそれぞれ撮像する範囲を調節できるように、矢印Bの方向に回動可能に構成されている。カメラ21及び31は動画を撮像可能な機能を有している。また、第1の本体20には、操作ボタン群24が設けられ、利用者により例えば「記録」ボタンが押されると、例えば会話支援装置1は、カメラ21等で撮像された画像を記憶する。

- [0030] また、会話支援装置1は、メインプロセッサ3、サブプロセッサ5及び6、RAM4及び7、主記憶部8、マイクロフォン23、音圧レベル判定部9、画像／音声再生処理部13

、スピーカ33、時刻生成部18を有する。

- [0031] メインプロセッサ3は、会話支援装置1を全体的に統括して制御し、また、サブプロセッサ5及び6のデータ処理のスケジュール管理を行う。サブプロセッサ5は、第1の本体20でのデータ処理を行い、サブプロセッサ6は、第2の本体30でのデータ処理を行う。各サブプロセッサ5及び6は独立してデータを処理することができるようになっている。RAM(Random Access Memory)4は、例えばメインプロセッサ3及びサブプロセッサ5の作業領域となり、RAM7は、サブプロセッサ6の作業領域となる。
- [0032] カメラ21及び31から取得されるそれぞれの画像フレームには、カメラ21及び31のどちらで撮像されたかを識別可能なようにカメラIDが付される。さらに、例えば、撮像された画像フレームには、該フレームごとに、かつ、取得された時間順に固有のシーケンスIDが付され、さらにタイムスタンプが付される。タイムスタンプは時刻生成部18により生成される。これにより、メインプロセッサ3、サブプロセッサ5及び6は、取得した画像フレームがどちらのカメラで撮像されたかを認識でき、また各画像フレームの順番やフレームごとの撮像時刻を把握することができる。また、カメラ21及び31で取得される画像フレームのフレームレートは例えば5〜20フレーム／秒とすることができる。
- [0033] また、マイクロフォンで取得される音声情報についても同様に、例えば所定の時間ごとにタイムスタンプが付される。この場合の所定の時間とは、上記フレームレートに合わせる事が好ましい。
- [0034] 図3は、主記憶部8に記憶されたソフトウェアを示す図である。主記憶部8は、例えばハードディスクや半導体メモリ等を用いることができる。主記憶部8は、画像解析プログラム11、音声解析プログラム12、画像／音声データベース17を格納している。画像解析プログラム11は、カメラ21で撮像された利用者の画像、特に口唇の動きを解析し文字情報または音声情報に変換する読唇機能、カメラ31で撮像された例えば利用者が行う手話の画像を解析し文字情報または音声情報に変換する手話画像解析機能等を有する。第2の本体30は、カメラ21で撮像された画像情報が変換された文字情報を表示する。また、第1の本体20は、カメラ31で撮像された手話の画像情報が変換された文字情報を表示する。画像解析プログラム11は、上記利用者の画

像を文字情報に変換するだけでなく手話情報に変換したり、手話の画像を音声情報に変換したりすることも可能となっている。

[0035] 音声解析プログラム12は、マイクロフォン23で集音された利用者である話者の声音を解析し文字情報等に変換するプログラムである。第2の本体30は、当該変換された文字情報を表示する。音声解析プログラム12は、当該文字情報だけでなく、手話情報に変換することも可能となっている。

[0036] 図4は、予め記憶された画像／音声データベース17の一例を示す図である。このデータベース17は、具体的には、人の口唇の画像パターン25、音響パターン26、文字情報27、手話情報28がそれぞれ対応づけられたテーブルである。

[0037] 口唇の画像パターン25は、例えば人が「あ」と発声したときの口唇やその口唇の近傍の画像である。口唇の画像だけでなく、その口唇の近傍の画像、例えば口唇の周囲のしわ等の情報をも用いることで、より高精度な画像認識が可能となる。例えば「え」という言葉と、「い」という言葉では、そのしわのより方が異なるからである。また、口唇画像パターン25は、1文字ごとではなく単語ごとに記憶されていることが好ましい。

[0038] 音響パターン26は、例えば人が発声したときの単語、語彙、あるいはセンテンスごとの音響特性である。音響特性とは、例えば周波数、音圧、ある単語と単語との時間間隔、ある語彙やセンテンスの総時間間隔等である。

[0039] 文字情報27は例えばテキスト情報である。手話情報28は、手話画像のパターンであり、例えば指文字または手話単語ごとに当該手話の画像パターンが記憶されている。

[0040] 図5は、図1に示す音圧レベル判定部9の構成を示すブロック図である。音圧レベル判定部9は、閾値設定部15及び比較部16を有する。音圧レベル判定部9は、例えばマイクロフォン23から入力された利用者の声音の音圧レベルが、閾値設定部15で設定された閾値以上になったと比較部16によって判断された場合に、カメラ21等から取り込まれる画像の解析を開始するための機能である。この場合、利用者が操作ボタン24を操作することで、閾値を設定できるようにしてもよい。また、音圧レベル判定部9は、音圧レベルではなく、騒音レベルで判定するようにしてもよい。

[0041] 図6は、第1の本体20と第2の本体30との接続部分を示す拡大図である。第1の本

体20と第2の本体30とは、それぞれの他端部2aと3bとで接続機構10により接続されている。接続機構10は、軸部10a及び10bを有している。軸部10bによって、第1の本体20と第2の本体30とが図1に示すようにAの方向に折りたたみ可能になっている。折りたたみの角度は、第1の表示部22及び第2の表示部32の表示面同士が対面するように、また当該表示面同士が互いに逆方向に向くような角度が可能である。すなわち角度Aが360度となるように、第1の本体20と第2の本体30とが接続機構10によって接続されている。また、軸部10aによって、第1の表示部22の表示面に対し第2の表示部32の表示面がねじれるようになっている。つまり、X軸を軸として第1の本体20と第2の本体30とが回転するようになっている。例えば、一方の表示部に対して他方の表示部が例えば270度回転する。このように、折りたたみ及びねじれが可能に構成されることで、会話する2人の位置が限定されなくなる。例えば2人が対面しても話すこともできるし、横に並んで話すこともできる。

[0042] 以上のように構成された会話支援装置1の動作について説明する。図7は、その動作を示すフローチャートである。

[0043] 例えば、健常者と聴覚障害者とが会話する場合について説明する。健常者は第1の表示部22を見ながら、聴覚障害者は第2の表示部32を見ながら会話する。

[0044] まず、会話支援装置1の図示しない電源が投入されると、カメラ21は健常者の画像の撮像を開始し、カメラ31は聴覚障害者の撮像を開始する(ステップ701)。この場合、撮像される動画は、RAM4等に所定の時間分、または所定の容量分記憶するようにすることができる。また、このときRAM4等に保存されていく画像フレームを、古い順に削除していくようにしてもよい。あるいは当該画像フレームを順に主記憶部8に保存するようにしてもよい。

[0045] 撮像が開始されると、メインプロセッサ3、またはサブプロセッサ5等は、カメラ21による撮像画像をリアルタイムで第2の表示部32に表示し、さらに、カメラ31による撮像画像をリアルタイムで第1の表示部22に表示する(ステップ702)。この場合、メインプロセッサ3の監視の下、サブプロセッサ5が第2の表示部32の表示を担当し、サブプロセッサ6が第1の表示部22の表示を担当するようにしてもよい。

[0046] 撮像を開始すると、会話支援装置1は、健常者または聴覚障害者の声音の入力待

ち状態となる(ステップ703)。入力があった場合(ステップ704のYES)、音圧レベル判定部9により、入力された声音の音圧レベルが閾値以上か否かが判定される(ステップ705)。入力された声音の音圧レベルが閾値以上である場合には、メインプロセッサ3は、画像解析プログラム11及び音声解析プログラム12を用いて、カメラ21等で撮像される画像の解析処理及びマイクロフォン23で取得される声音の解析処理を開始する。この場合、メインプロセッサ3の解析処理命令により、各サブプロセッサ5及び6が分担して解析処理を行う。この場合、例えば、サブプロセッサ5が画像解析を行うようにし、サブプロセッサ6が音声解析を行うようにすればよい。

[0047] このように解析処理された後は、メインプロセッサ3等は、所定の時間だけステップ705のYESの状態を保持する。すなわち、最初に健常者が話した言葉が途切れたりしても、そのままステップ706を実行し続ける。所定の時間とは、例えば1秒、あるいは数秒程度である。

[0048] 画像解析プログラム11により、特に健常者の口唇の画像解析が行われる。具体的には、人間の口の位置は顔の下部と決まっているため、このことを予め会話支援装置1が把握しておけば、健常者の顔全体をスキャンし顔の輪郭を抽出することで口唇の位置を把握することができる。画像解析処理のフローを図8に示す。カメラ21で取得し続けている健常者の口唇の画像と、主記憶部8に記憶された各口唇画像パターン25とを随時比較していく(ステップ801)。そして、対応する口唇画像パターンを選択し(ステップ802)、選択された画像パターンを抽出する(ステップ803)。

[0049] ここで、画像解析の手法としては、具体的には、白黒で2値化された数値のマトリクスで画像を認識する。より詳しくは、例えば、2値化された画像情報の、各画素の数値が一致する数が最も多い画像フレームを、撮像した口唇の画像に対応する一の口唇画像パターンとみなして扱えばよい。口唇の形状や手の形状(手の形状を撮像する場合は、手話の画像を解析する場合)のみを把握すればよいので、最低限このような2値化データで足りる。これにより主記憶部8またはRAM4等の節約にもなる。

[0050] サブプロセッサ5等は、抽出された一の画像パターンを基に、この一の画像パターンに対応する手話情報28を抽出して、第2の表示部32に表示する(ステップ707)。このとき第2の表示部32の表示例を図9に示す。この例では、例えば「こんにちは」と

いう情報が手話の画像で表示されている。さらに、サブプロセッサ5等は、抽出された一の画像パターンを基に、この一の画像パターンに対応する文字情報27を抽出して、図に示すように第2の表示部32に表示するようにしてもよい。聴覚障害者は、第2の表示部32を見ることにより、その内容を理解することができる。

- [0051] また、音声解析プログラム12により、健常者が発声する音声の解析が行われる。具体的には、例えばマイクロフォン23で取得し続けている声音情報と、主記憶部8に記憶された音響パターン26とを随時比較していく。そして、対応する音響パターン26を選択し、選択された音響パターンを抽出する。この音声の解析処理は、例えば隠れマルコフモデルを用いることによって行われる。抽出された音響パターンに対応する手話情報または文字情報を抽出し、これを第2の表示部32に表示する。また、音声解析処理では、取得した音に含まれる雑音の除去、ひずみの補正等も行われる。
- [0052] 上記ステップ802では、例えば、音声解析プログラム12による音声解析処理と同期して処理が行われる。具体的には、例えば取得された画像フレームと同じタイムスタンプが付された部分の音声データの解析が行われ、画像解析処理によってステップ707で抽出された手話情報と、音声解析処理によってステップ707で抽出された手話情報とが一致するか否かが判断される。ここで、一致しない場合は、再度の画像解析処理または音声解析処理が行われる。あるいは、一般的に音声認識の方が認識率が高いので、音声解析処理で抽出された方を優先して用いてもよい。このように、画像と音声とにより解析処理が行われることにより、健常者の話す内容の認識精度が向上する。
- [0053] また、ステップ707において、サブプロセッサ6等は、サブプロセッサ5が第2の表示部32に、抽出された手話情報または文字情報を表示する時間とほぼ同期して、当該手話情報または文字情報を第1の表示部22に表示するようにしてもよい。これにより、健常者は、第1の表示部22を見ながら自身が話した内容について、変換ミス等がないか等を確認することができる。
- [0054] 一方、聴覚障害者が話す場合、サブプロセッサ6等は、カメラ31で取得された聴覚障害者が行っている手話の画像を解析処理し、その手話画像に対応する手話情報28を抽出し、さらにその手話情報に対応する音響パターン26を抽出する。そして、こ

の音響パターン26を基に音声情報に変換し、スピーカ33に出力する。あるいは、サブプロセッサ6等は、上記抽出された手話情報28に対応する文字情報27を抽出し、これを第1の表示部22に表示する。

[0055] 以上のように、本実施の形態では、健常者が第1の表示部22を見ながら、聴覚障害者が第2の表示部32を見ながら、会話することができる。したがって両者は、効率よくスムーズに会話することができる。

[0056] 本実施の形態では、音圧レベル判定部9を設ける構成としたので、従来のように、話し始めるたびに画像解析や音声認識の制御キーのようなものを押す必要がなくなり、会話をスムーズにすることができる。

[0057] 次に、本発明の他の実施の形態について説明する。この実施の形態において、上記実施の形態に係る会話支援装置1の部材、機能、動作と同様なものについては、その説明を簡略または省略し、異なる点を中心に説明する。

[0058] 図10は、例えば複数の人の口唇の形状パターンと、それらの人ごとの変換テーブルのIDとが対応付けられたテーブルを示す。このテーブル40は、例えば主記憶部8(図2参照)に記憶されていればよい。図11に変換テーブルを示す。この変換テーブル45は、図4で示した口唇の画像パターン25と文字情報27との対応関係と同様なものである。また、この変換テーブル45は、当該複数の人のうちのある1人(例えば、変換テーブルID=7)の固有の変換テーブルである。なお、図11では、一文字ずつの対応関係を示しているが、図4で説明したように、もちろん単語、語彙、センテンスごとの変換テーブルであってもかまわない。複数の人とは、例えば5人、10人、20人、あるいはそれ以上であってもよい。

[0059] 図10に示す口唇形状パターンは、例えば、撮像されたものが予めデフォルトで記憶されていてもよいし、利用者が収集しても記憶してもかまわない。また、口唇形状パターンは、口が閉じられた状態(何も発声していない状態)でもよいし、ある言葉「あ」とか、「こんにちは」とか何らかの言葉を発声している状態であってもよい。「あ」なら、静止画でよいし、「こんにちは」なら動画とすればよい。

[0060] 図12は、本実施の形態に係る会話支援装置の動作の一部を示すフローチャートである。

- [0061] この会話支援装置を用いて会話する前に、まず、初期設定を行う。この初期設定の動作について図12を参照しながら説明する。例えば、会話する者が健常者と聴覚障害者である場合、健常者の口唇をカメラ21等で撮像する(ステップ201)。この場合、例えば操作ボタン群24に「初期設定」ボタンを設け、健常者により初期設定ボタンが押されることによって初期設定が開始されるようにしてもよい。また、健常者は、口が閉じられた状態(何も発声していない状態)でもよいし、ある言葉「あ」とか、「こんにちは」とか何らかの言葉を発声している状態をカメラ21で撮像するようにしてもよい。
- [0062] メインプロセッサ3、サブプロセッサ5または6等は、撮像された健常者の口唇をRAM4等に記憶し、上記画像解析プログラム11を用いて、健常者の口唇の形状と、主記憶部8に記憶された上記複数の人ごとの口唇形状パターンとを比較していく(ステップ202)。そして、上記複数の人ごとの口唇形状パターンのうち、健常者の口唇の形状と最も類似する一の形状パターンを選択する(ステップ203)。具体的には、例えば、白黒で2値化された画像情報の、各画素の数値が一致する数が最も多い画像フレームを選択すればよい。ステップ203の後、メインプロセッサ3等は、選択された形状パターンに係る変換テーブル45を抽出する(ステップ204)。その後は、メインプロセッサ3等はテーブル40を参照しながら図7で示した動作を行えばよい。これにより、より高精度に利用者の画像認識を行うことができる、また、より高精度に手話情報等への変換処理を行うことができる。したがって、変換処理の変換ミス等の発生率を低減することができ、スムーズに会話することができる。
- [0063] 本発明は以上説明した実施の形態には限定されるものではなく、種々の変形が可能である。
- [0064] 例えば、上記の実施の形態では、健常者と聴覚障害者との会話を例に示したが、例えば日本人と外国人との会話であってもよい。また、本実施の形態によれば、聴覚障害者と視覚障害者との会話を成立させることができる。この場合、視覚障害者の発声した音を、文字情報または手話情報に変換して表示部に表示すれば、聴覚障害者がそれを見て理解することができる。一方、聴覚障害者が行う手話情報を音声情報に変換して、スピーカ33で出力すれば、視覚障害者はその出力された音声を聞くことにより会話することができる。

- [0065] 上記実施の形態では、健常者が話した内容が図9に示したように第2の表示部32に手話情報または文字情報として表示される例を示した。しかし、手話情報等の代わりに、または手話情報等に加えて、健常者の口唇の動画像が第2の表示部32に表示されるようにしてもよい。この場合、第2の本体30で表示される口唇の画像はアニメーション画像に変換されて表示されてもよい。
- [0066] 会話支援装置1は、図1で示した形態に限られない。例えば図13に示すような会話支援装置を用いるようにしてもよい。この会話支援装置50は、例えばラップトップ型のコンピュータ程度の大きさである。この会話支援装置50は、第1の本体60と第2の本体70とが接続部51で接続され、任意の角度Cで折りたたみ可能に構成されている。そして第1の本体60及び第2の本体70にそれぞれ任意の角度Dで回動可能な脚部69及び79が設けられている。これらの脚部69及び70に、例えば図示しない操作ボタン、スピーカ、またはマイクロフォン等が装備されるようにしてもよい。符号61及び71はカメラであり、符号62及び72は表示部(表示パネル)である。これにより、例えば健常者34と、聴覚障害者35とが会話することができる。
- [0067] 上記実施の形態では、健常者が最初に発声して音圧レベル判定部9により、入力された音の音圧レベルを判定する例を説明したが、聴覚障害者が最初に発声した音の音圧レベルを判定するようにしてもよい。聴覚障害者が話すときの発声音は、聞き手にとってはわかりにくい場合もある。しかし、聴覚障害者が何らかの言葉を発声しさえすれば、その音圧レベルを判定することにより、そのときに聴覚障害者が行っている手話の画像の解析処理を開始することができる。
- [0068] 利用者が操作ボタン群24を操作することにより、表示部20等に表示される画像や、スピーカから出力される音声等のトリックプレイ操作を行うことができる構成としてもよい。トリックプレイ操作とは、例えば早送り、巻戻し等である。例えば、画像の早送りの場合、RAM上のデータに格納された画像フレームをシーケンス番号の小さい順番に、3, 6, 9, ...等のように間欠的に表示部に表示させるようにすればよい。このようなトリックプレイは、画像／音声再生処理部13(図2参照)で行うことができる。
- [0069] 図11で示した変換テーブル45は、口唇画像パターンと文字情報との変換テーブルであった。しかし、口唇画像パターンと手話情報との変換テーブル、または、口唇

画像パターンと音声情報との変換テーブルであってもよい。

- [0070] 図12においては、利用者が会話を始める前に初期設定が行われる例を説明した。しかし、初期設定しなくても、会話の開始された直後から、図12に示すような動作が行われるようにしてもよい。
- [0071] 上記実施の形態では、第1の本体20のみにマイクロフォン23が設けられる構成とした。しかし、マイクロフォンが第2の本体30にも設けられるようにしてもよい。また、第2の本体30のみにスピーカ33が設けられるようにしたが、第1の本体20にも設けられていてもかまわない。
- [0072] サブプロセッサの数は2つに限られず、1つであってもよいし、3つ以上であってもよい。

図面の簡単な説明

- [0073] [図1]本発明の一実施の形態に係る会話支援装置を示す斜視図である。
[図2]会話支援装置の構成を示すブロック図である。
[図3]主記憶部8に記憶されたソフトウェアを示す図である。
[図4]画像／音声データベースの一例を示す図である。
[図5]音圧レベル判定部の構成を示すブロック図である。
[図6]第1の本体と第2の本体との接続部分を示す拡大図である。
[図7]会話支援装置の動作を示すフローチャートである。
[図8]図7に示すステップ706の画像解析処理のフローチャートである。
[図9]変換処理後の第2の表示部32に表示される画像の一例を示す図である。
[図10]複数の人の口唇の形状パターンと、それらの人ごとの変換テーブルのIDとが対応付けられたテーブルを示す図である。
[図11]変換テーブルを示す図である。
[図12]他の実施の形態に係る会話支援装置の動作の一部(初期設定)を示すフローチャートである。
[図13]本発明のさらに別の実施の形態に係る会話支援装置を示す側面図である。

符号の説明

- [0074] 1, 50…会話支援装置

2a, 3a…他端部
3…メインプロセッサ
4, 7…RAM
5, 6…サブプロセッサ
8…主記憶部
9…音圧レベル判定部
11…画像解析プログラム
12…音声解析プログラム
13…音声再生処理部
14、33…スピーカ
15…閾値設定部
16…比較部
17…画像／音声データベース
20, 60…第1の本体
21、31、61、71…カメラ
23…マイクロフォン
25…口唇画像パターン
26…音響パターン
27…文字情報
28…手話情報
30, 70…第2の本体

請求の範囲

- [1] 第1の利用者及び第2の利用者を撮像する手段と、
前記撮像された前記第1の利用者の第1の画像情報を前記第2の利用者が理解可能な第1の情報に変換する第1の変換手段と、
前記撮像された前記第2の利用者の第2の画像情報を前記第1の利用者が理解可能な第2の情報に変換する第2の変換手段と、
第1の表示面を有し、前記変換された前記第1の情報を前記第1の表示面に表示する第1の表示手段と、
前記第1の表示面とは異なる角度に配置可能な第2の表示面を有し、前記変換された前記第2の情報を前記第2の表示面に表示する第2の表示手段と
を具備することを特徴とする会話支援装置。
- [2] 請求項1に記載の会話支援装置であって、
前記第1の利用者の声音を集音する手段と、
前記集音された声音を前記第1の情報に変換する第3の変換手段と
をさらに具備することを特徴とする会話支援装置。
- [3] 請求項1に記載の会話支援装置であって、
前記第1の変換手段は、前記第1の画像情報を、前記第1の情報として文字情報及び手話情報のうちいずれか一方に変換する手段を有することを特徴とする会話支援装置。
- [4] 請求項1に記載の会話支援装置であって、
前記第1の変換手段は、前記第1の画像情報を、前記第1の情報として音声情報に変換する手段を有し、
当該会話支援装置は、前記変換された音声情報を出力する手段をさらに具備することを特徴とする会話支援装置。
- [5] 請求項3に記載の会話支援装置であって、
前記第1の表示手段は、前記第1の表示面に前記第2の情報を表示する手段を有することを特徴とする会話支援装置。
- [6] 請求項5に記載の会話支援装置であって、

前記第1の表示手段は、前記第2の表示手段が前記第2の情報を表示するタイミングとほぼ同期して該第2の情報を表示する手段を有することを特徴とする会話支援装置。

- [7] 請求項3に記載の会話支援装置であって、
前記第1の変換手段が、前記第1の画像情報を文字情報及び手話情報に変換する場合、
前記第1の表示手段は、前記第1の変換手段で変換された前記手話情報を前記第1の表示面に表示し、
前記第2の表示手段は、前記第1の表示手段が前記手話情報を表示するタイミングとほぼ同期して、前記第1の変換手段で変換された前記文字情報を表示する手段を有することを特徴とする会話支援装置。
- [8] 請求項2に記載の会話支援装置であって、
前記第1の利用者の声音の音圧レベルまたは騒音レベルを計測する手段と、
前記音圧レベルまたは騒音レベルの閾値を設定する手段と、
前記測定された音圧レベルまたは騒音レベルが前記閾値以上である場合に、前記第3の変換手段による変換処理を実行する手段と
をさらに具備することを特徴とする会話支援装置。
- [9] 請求項2に記載の会話支援装置であって、
前記第1の利用者の声音の音圧レベルまたは騒音レベルを計測する手段と、
前記音圧レベルまたは騒音レベルの閾値を設定する手段と、
前記測定された音圧レベルまたは騒音レベルが前記閾値以上である場合に、前記第1の変換手段による変換処理を実行する手段と
をさらに具備することを特徴とする会話支援装置。
- [10] 請求項3に記載の会話支援装置であって、
人が発声したときの該人の口唇の複数の画像パターンと、該人が発声したときの音に対応する前記文字情報とを対応付けて記憶する手段をさらに具備し、
前記撮像手段は、前記第1の画像情報として第1の利用者の口唇の画像を撮像し、

前記第1の変換手段は、
前記撮像された口唇画像情報と、前記各画像パターンとを比較する手段と、
比較の結果、該各画像パターンのうち前記撮像された口唇画像情報に対応する一の画像パターンを選択する手段と、
選択された前記一の画像パターンに対応する前記文字情報を抽出する手段と
を有することを特徴とする会話支援装置。

- [11] 請求項10に記載の会話支援装置であって、
前記記憶手段は、
複数の人ごとに、前記画像パターンと前記文字情報とを対応付けて記憶する手段と、
前記各人の口唇の形状パターンを記憶する手段とを有し、
当該会話支援装置は、
前記撮像された口唇画像情報と、前記各形状パターンを比較する手段と、
比較の結果、前記各形状パターンのうち、前記撮像された口唇画像情報に最も類似する一の形状パターンを選択する手段と
をさらに具備することを特徴とする会話支援装置。

- [12] 請求項3に記載の会話支援装置であって、
前記人が発声したときの該人の口唇の複数の画像パターンと、該人が発声したときの音に対応する前記手話情報とを対応付けて記憶する手段をさらに具備し、
前記撮像手段は、前記第1の画像情報として第1の利用者の口唇の画像を撮像し、
前記第1の変換手段は、
前記撮像された口唇画像情報と、前記各画像パターンとを比較する手段と、
比較の結果、前記撮像された口唇画像情報に対応する該各画像パターンのうちの画像パターンを選択する手段と、
選択された前記一の画像パターンに対応する前記手話情報を抽出する手段と
を有することを特徴とする会話支援装置。

- [13] 請求項12に記載の会話支援装置であって、

前記記憶手段は、
複数の人ごとに、前記画像パターンと前記手話情報とを対応付けて記憶する手段と、
前記各人の口唇の形状パターンを記憶する手段とを有し、
当該会話支援装置は、
前記撮像された口唇画像情報と、前記各形状パターンを比較する手段と、
比較の結果、前記各形状パターンのうち、前記撮像された口唇画像情報に最も類似する一の形状パターンを選択する手段と
をさらに具備することを特徴とする会話支援装置。

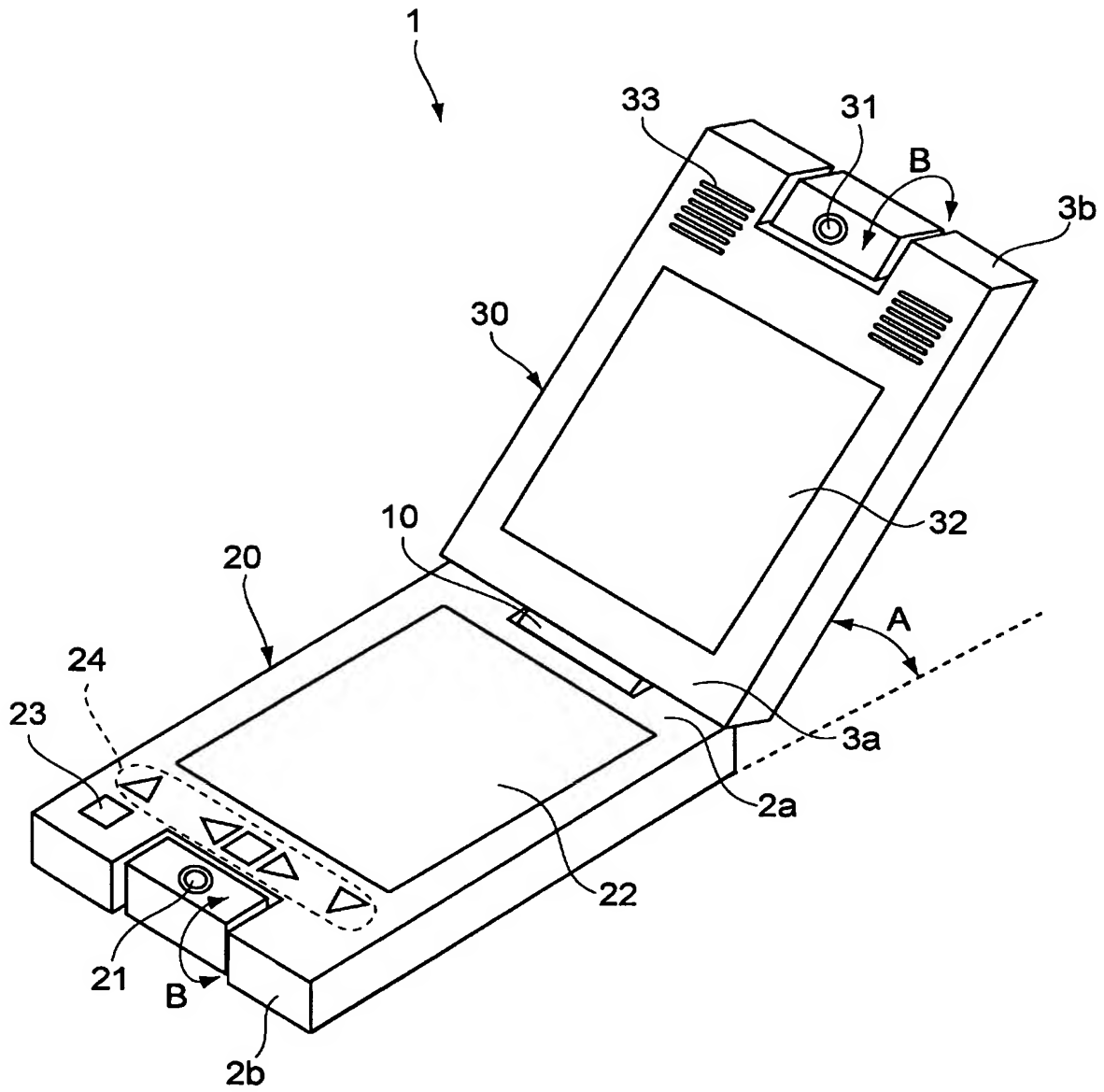
- [14] 請求項4に記載の会話支援装置であって、
前記人が発声したときの該人の口唇の複数の画像パターンと、該人が発声したときの音に対応する前記音声情報とを対応付けて記憶する手段をさらに具備し、
前記撮像手段は、前記第1の画像情報として第1の利用者の口唇の画像を撮像し、
前記第1の変換手段は、
前記撮像された口唇画像情報と、前記各画像パターンとを比較する手段と、
比較の結果、前記撮像された口唇画像情報に対応する該各画像パターンのうちの画像パターンを選択する手段と、
選択された前記一の画像パターンに対応する前記音声情報を抽出する手段と
を有することを特徴とする会話支援装置。

- [15] 請求項14に記載の会話支援装置であって、
前記記憶手段は、
複数の人ごとに、前記画像パターンと前記手話情報とを対応付けて記憶する手段と、
前記各人の口唇の形状パターンを記憶する手段とを有し、
当該会話支援装置は、
前記撮像された口唇画像情報と、前記各形状パターンを比較する手段と、
比較の結果、前記各形状パターンのうち、前記撮像された口唇画像情報に最も類

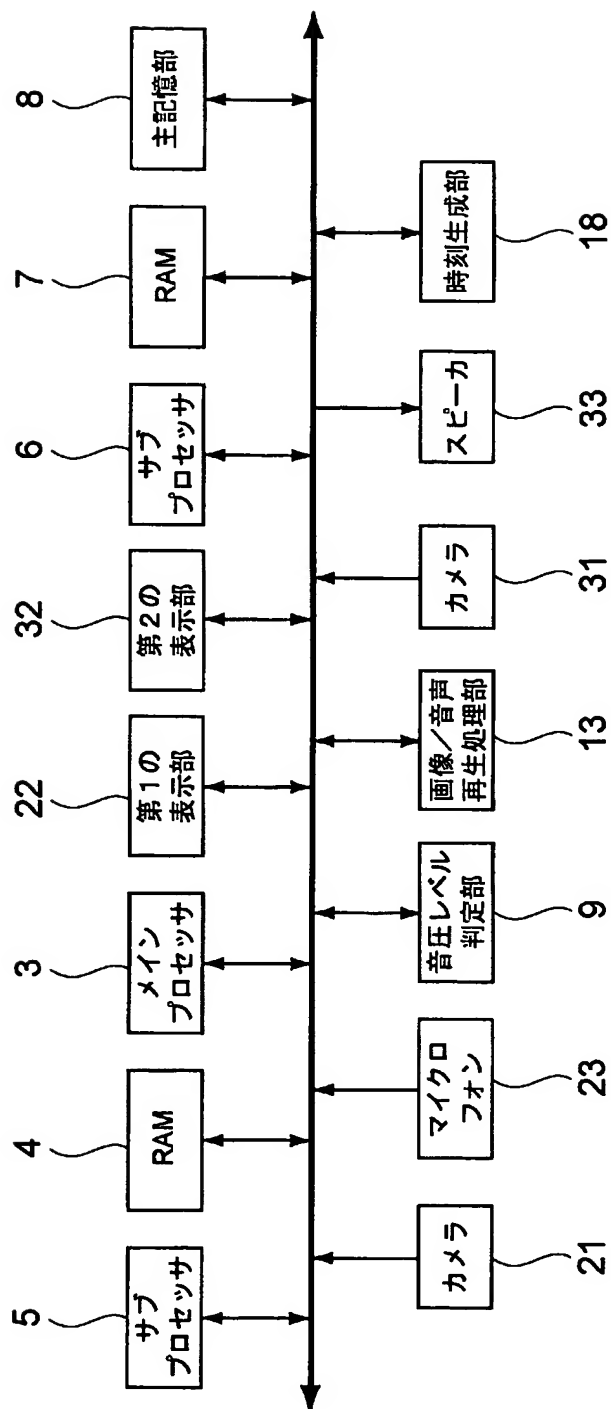
似する一の形状パターンを選択する手段と
をさらに具備することを特徴とする会話支援装置。

- [16] 請求項1に記載の会話支援装置であって、
前記第1の表示手段は、第1の端部を有し前記第1の表示面が配置される第1の部分
を有し、
前記第2の表示手段は、第2の端部を有し前記第2の表示面が配置され、前記第1
の表示面に対し前記第2の表示面をねじることができるように、かつ、前記第1の部分
に対し折りたたみできるように前記第1の端部と前記第2の端部とが接続された第2の
部分を有することを特徴とする会話支援装置。
- [17] 請求項1に記載の会話支援装置であって、
前記第1の情報を記憶する手段と、
前記記憶された第1の情報を前記第1の表示面に再生する手段と、
前記記憶された第1の情報をトリックプレイ操作する手段と
をさらに具備することを特徴とする会話支援装置。
- [18] 第1の利用者及び第2の利用者を撮像するステップと、
前記撮像された前記第1の利用者の第1の画像情報を前記第2の利用者が理解可
能な第1の情報に変換するステップと、
前記撮像された前記第2の利用者の第2の画像情報を前記第1の利用者が理解可
能な第2の情報に変換するステップと、
前記変換された前記第1の情報を前記第1の表示面に表示するステップと、
前記変換された前記第2の情報を、前記第1の表示面とは異なる角度に配置可能
な第2の表示面を表示するステップと
を具備することを特徴とする会話支援方法。

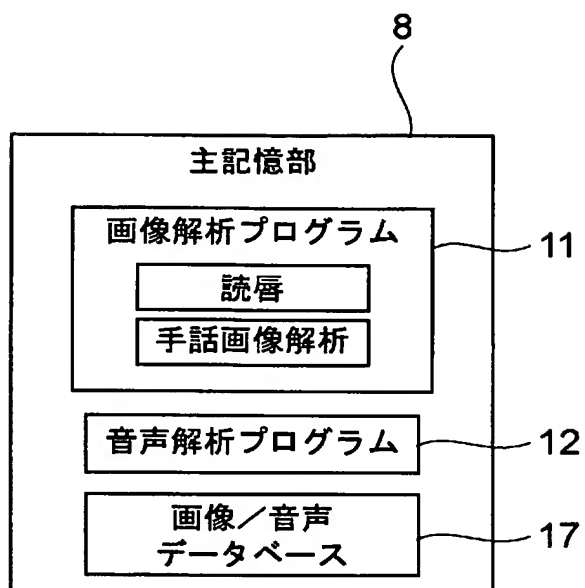
[図1]



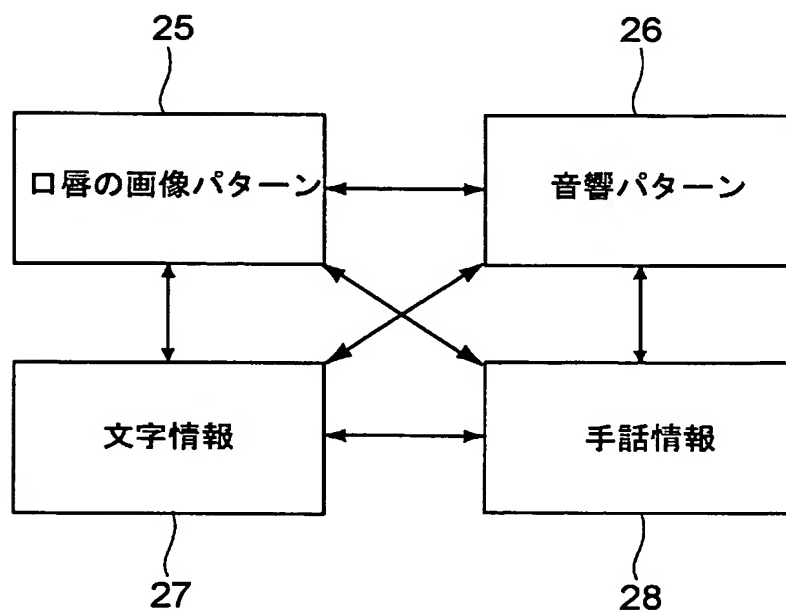
[図2]



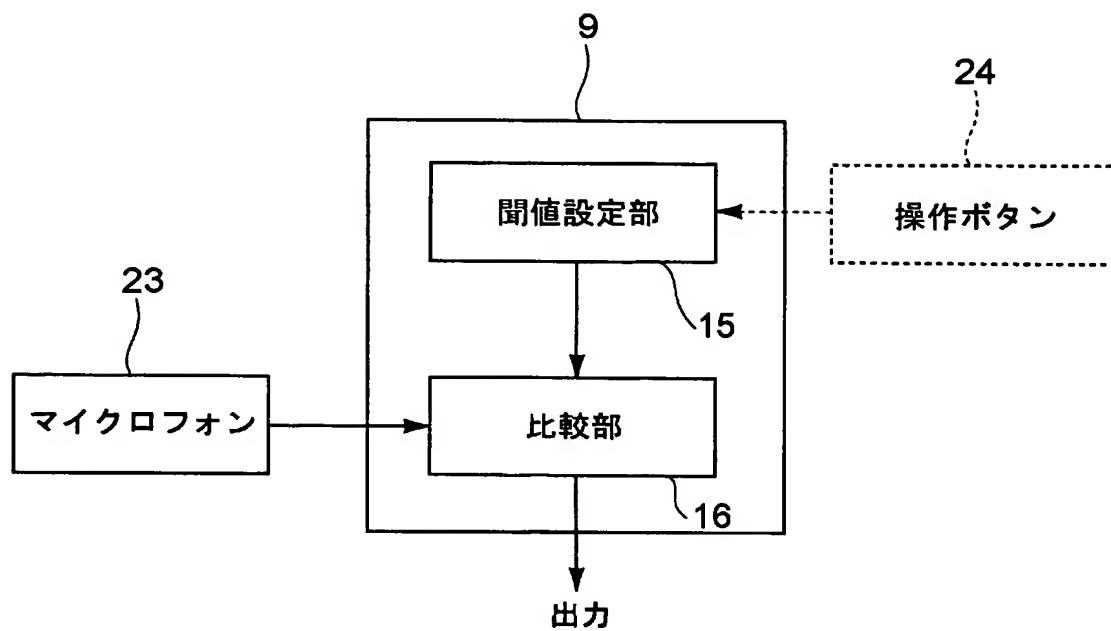
[図3]



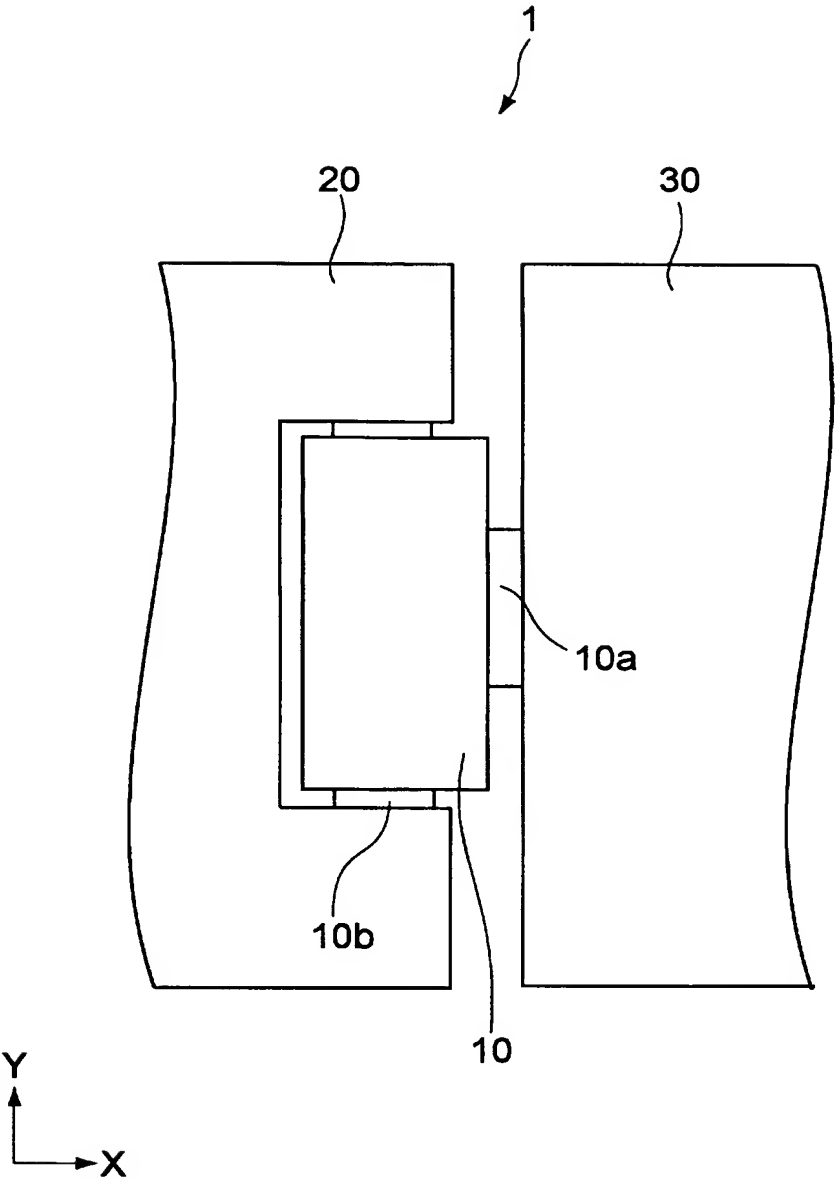
[図4]



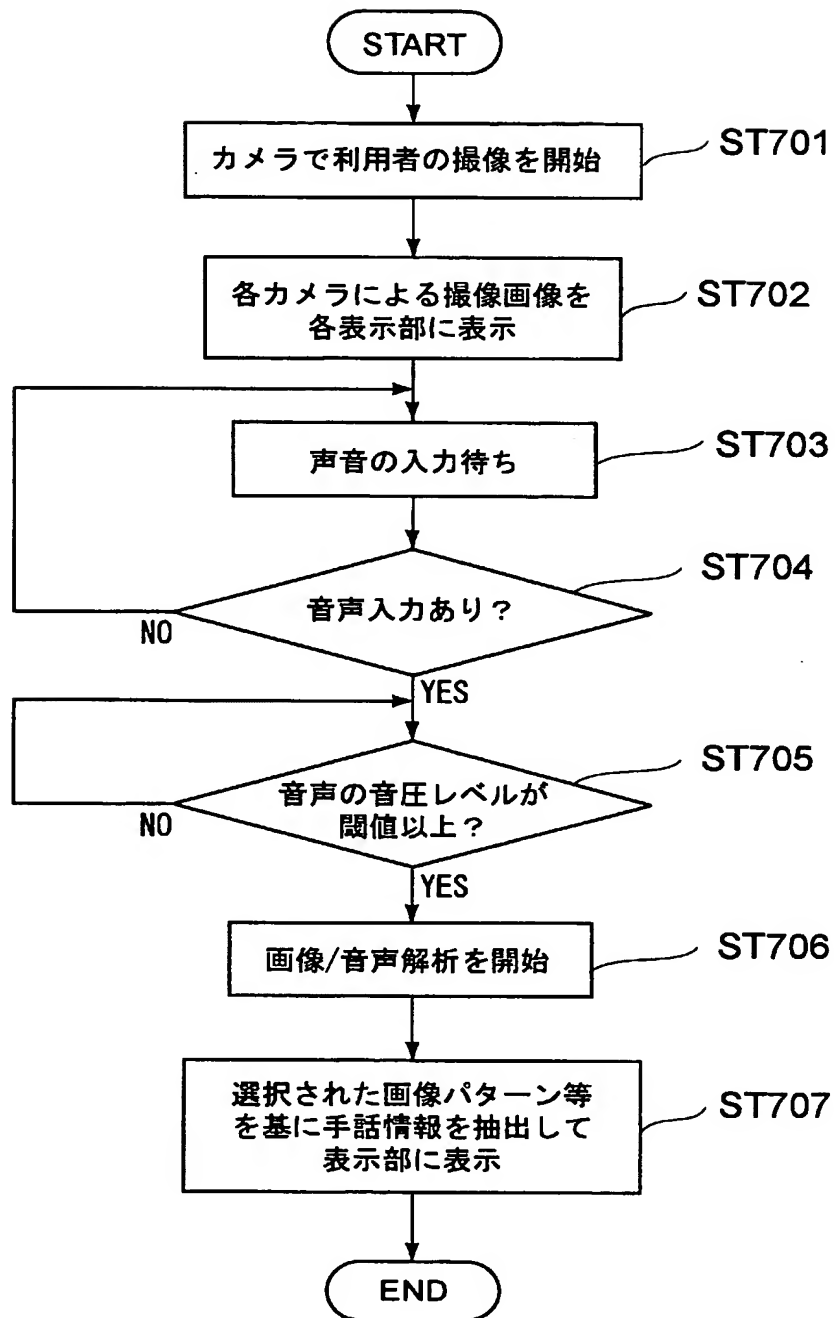
[図5]



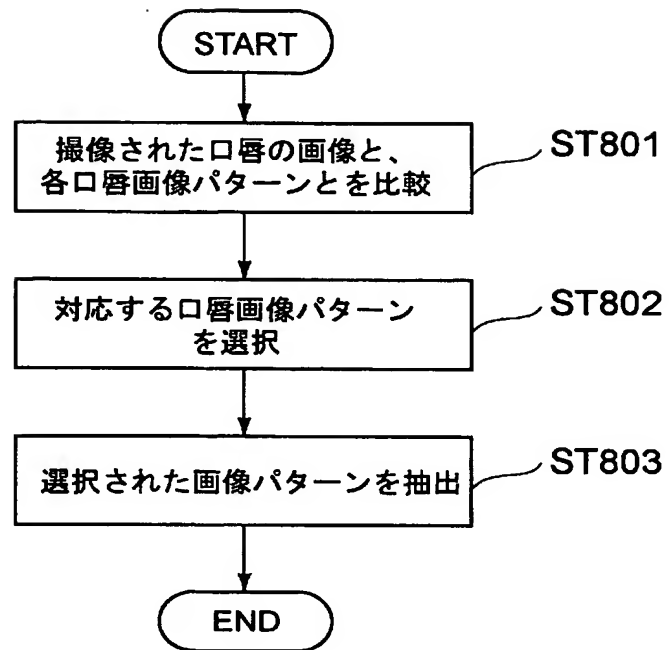
[図6]



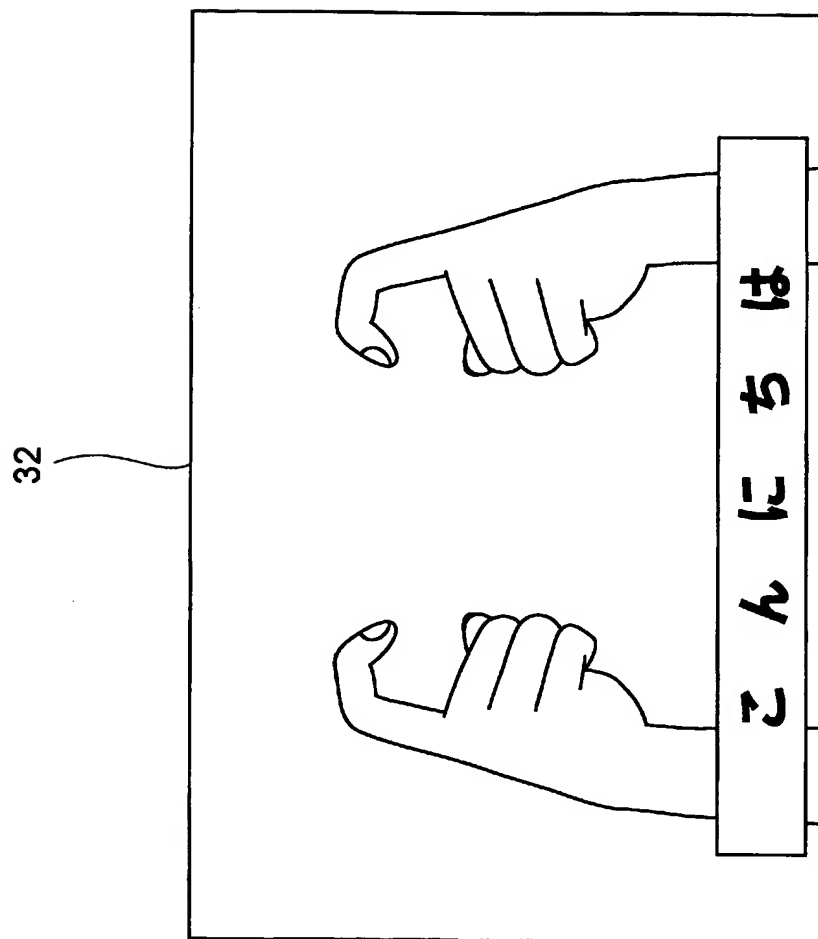
[図7]



[図8]






[図9]



[図10]




40
↓

| 口唇の形状パターン | 変換テーブルID |
|---|----------|
|  | 1 |
|  | 2 |
|  | 3 |
| ⋮ | ⋮ |

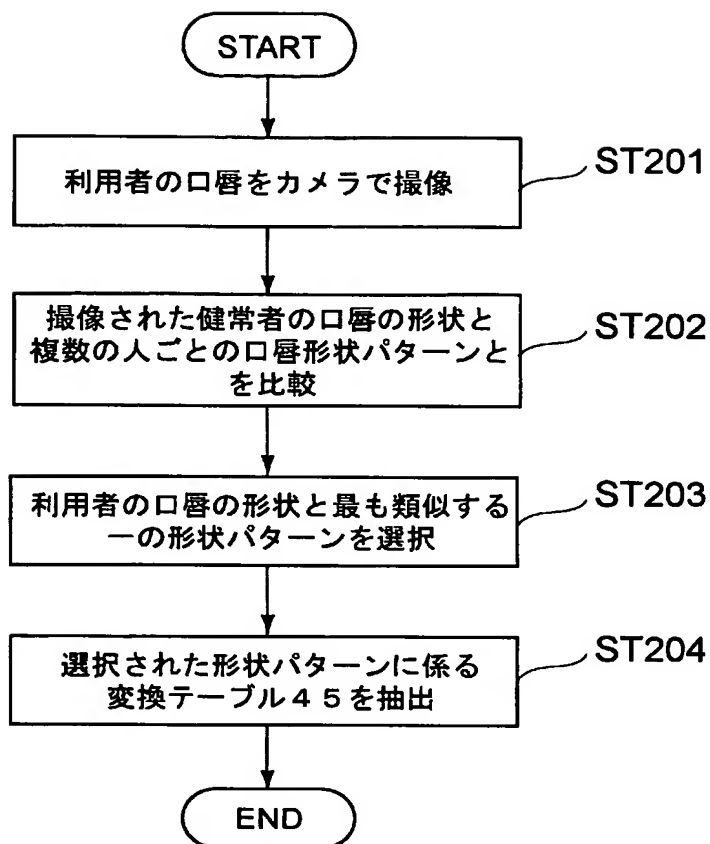
[図11]

45
↓

変換テーブルID=7

| 口唇の画像パターン | 文字情報 |
|---|------|
|  | あ |
|  | い |
|  | う |
| ⋮ | ⋮ |

[図12]



[図13]

